DOCUMENT RESUME

ED 432 042 EA 029 905

AUTHOR Chiu, Shuwan; Wardrop, James L.; Ryan, Katherine E.
TITLE Use of the Unbalanced Nested ANOVA To Examine the Relationship of Class Size to Student Ratings of Instructional Quality.
PUB DATE 1999-04-00
NOTE 55p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Class Size; Educational Quality; Elementary Secondary Education; *Student Attitudes; *Student Motivation; Teacher Effectiveness; Teacher Student Ratio
IDENTIFIERS *Instructor and Course Evaluation System

ABSTRACT

This report examines how class size and course motivation are related to evaluation results, its purpose being to provide a comparison of rating differences on the Instructor and Course Evaluation System (ICES). ICES offers faculty a computer-based catalog of over 1,000 survey items. The article is based on a study that sought to extend the ICES research by using an unbalanced nested ANOVA to analyze the effects of five factors: course motivation, course level, class size, academic discipline, and individual class differences on (1) student ratings of instructors; and (2) student ratings of overall course quality. The results of the ANOVA showed that course motivation, course level, class size, discipline, and individual class do influence student ratings of both teaching effectiveness and course quality. Different combinations of course motivation, level, and discipline displayed different patterns of student ratings. The significant two-way interactions (in which one of the factors alters the effect of the other) were course motivation and discipline, course level and class size, and course level and discipline. Generally, ratings of elective courses were higher than those of mixed courses, and ratings of mixed courses were higher than those of required courses. Higher level classes had higher ratings than lower level ones. Contains 27 references, 10 figures, and 9 tables. (RJM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EA

Use of the Unbalanced Nested ANOVA to Examine the Relationship of Class Size to Student Ratings of Instructional Quality

ED 432 042

Shuwan Chiu

James L. Wardrop

Katherine E. Ryan

University of Illinois at Urbana-Champaign

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Chiu

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

For presentation at the Annual Meeting of the American Educational Research Association, Montreal, Canada, April, 1999

EA029905

BEST COPY AVAILABLE

Summary

This study utilized a 5-way unbalanced nested ANOVA model to evaluate student ratings of teaching. An effective sample included three types of course motivation, four course levels, seven levels of class size, 9 types of academic discipline, and about 3,754 classes with a total of 106,303 students. The dependent variables were ratings for two global items: one involved teaching effectiveness of the instructor, the other involved the overall quality of the course. Based on the ANOVA results, the distribution for each class size was developed. The distributions were used to compare the evaluation of teaching among students in classes of the same size and course motivation.

Introduction

In order to improve the quality of courses and the effectiveness of teaching, student evaluation of teaching is very important. Among the many methods of determining effective teaching in higher education, student ratings are commonly used because of their reliability and validity (Howard, Conway, & Maxwell, 1985; Marsh, 1984, 1987; Murray & Renaud, 1995; Murray, Rushton, & Paunonen, 1990; Newport, 1996).

Typically, student ratings are obtained through evaluation questionnaires completed at the end of each term. The evaluation questionnaires vary in format and type. Some colleges and universities use locally developed evaluation forms; some purchase commercial instruments such as the Instructional Development and Effectiveness Assessment, or IDEA (Cashin & Downey, 1992), developed at Kansas State University. Others use assessment forms developed on university campuses.

Braskamp and Ory (1994) are among those who conducted studies of student ratings of instructors and courses obtained from evaluation questionnaires on the Instructor and Course Evaluation System (ICES). Factors studied have included gender of students and/or instructors; teaching experience; course level; class size; and the personality of the instructor, among others.

Many studies have examined the variables of course motivation, class levels and disciplines. Generally, the ratings of elective courses are higher than those of required courses; the ratings of higher level classes are higher than those of lower level classes (Aleamoni & Graham, 1974; Feldman, 1978; Kulik & McKeachie, 1975; Marsh, 1984, 1987; Wigington, Tollefson & Rodriguez, 1989); the ratings of courses

in the fine arts and humanities are higher than those in the science, engineering and math-related (Cashin & Downey, 1995; Feldman, 1978, 1987; Franklin & Theall, 1995; Kulik & McKeachie, 1975; Marsh, 1984, 1987; Murray & Renaud, 1995).

Some faculty members have claimed that there are meaningful rating differences between large and small classes. There are many studies which found that the student ratings of their instructors in small classes are higher than those in large classes (Bausell & Bausell, 1979; Feldman, 1978, 1987; Kulik & McKeachie, 1975; Marsh, 1984, 1987; Wigington, Tollefson & Rodriguez, 1989). Feldman (1978, 1987) also examined how course motivation (required or elective) influenced student ratings. The focus of this study is on how class size and course motivation are related to evaluation results, and this approach is intended to yield a comparison of rating differences on the ICES.

The ICES developed at the University of Illinois offers faculty a computer-based catalogue of over 1,000 survey items. The current ICES feedback report places instructors in one of four percentage categories of ratings by course motivation (where course motivation means whether the course is required, mixed or elective): the first 20% immediately following the lowest score, middle 40%, next 20%, top 10%.

This study seeks to extend the ICES research by using an unbalanced nested ANOVA to analyze the effects of five factors: course motivation (required/ elective), course level, class size, academic discipline, and individual class differences on (1) student ratings of instructors and (2) student ratings of overall course quality.

The Importance of Student Ratings

In higher education, there are many methods to determine the effectiveness of teaching, including ratings by students, colleagues, administrators, and teachers themselves (Kulik & McKeachie, 1975; Marsh, 1984). Many administrators use student ratings as a consideration in determining instructors' salary, tenure, and promotion (Feldman, 1979; Haladyna & Hess, 1994; Kulik & McKeachie, 1975; Marsh, 1984; Newport, 1996).

There have been many studies demonstrating the reliability and validity of student ratings as compared to other types of teaching evaluation (Howard, Conway, & Maxwell, 1985; Marsh, 1984; Murray & Renaud, 1995; Murray, Rushton, & Paunonen, 1990; Newport, 1996). Students use a variety of evidence to make decisions on their ratings (Benz & Blatt, 1996). It has been shown that student ratings are considered to be among the most effective and accurate methods of teaching evaluations (Howard, Conway, & Maxwell, 1985). Besides, student ratings are a good way to evaluate the quality of teaching performance because the students are the audience of the classroom. (Kulik & McKeachie, 1975).

Indeed, the student evaluations of teaching provide essential and useful feedback for instructors (Simpson, 1995). The instructors' responses to this feedback can contribute to the improvement and effectiveness of teaching. Therefore, since student evaluations are so useful, a brief discussion of the following major factors affecting student ratings of their instructors is given below: size of the class, course motivation, class level, and academic discipline.

In addition, it is worth noting that there is a positive correlation between student ratings and academic achievement (Simpson, 1995). Therefore, the course grade could be referred to as another of the primary factors that affect the ratings of the teacher by the students.

In the opinion of Feldman (1978), different data or different statistical methods would get different results. To search for a better way to analyze the results of student ratings of teaching evaluation is very important.

Based on the findings of Marsh (1984, 1987), the separated statistical models for evaluation items are highly recommended. Owing to the types of categories for the effects and the association between effects, analysis of variance is more suitable than a variety of regression models and correlation matrices.

For the relationship between student ratings and the effect of class size, most studies presented a nonlinear relationship. As to the association between course level and student ratings, most of the studies found a positive correlation between them.

My study seeks to extend the earlier research by using an unbalanced nested ANOVA to analyze characteristics that might affect student ratings of instructors and student ratings of overall course quality. The effects of course motivation, class size, course level, and discipline are very important. Those effects will be taken into consideration.

Method

Data Source

This study was conducted using data from ICES questionnaires. An ICES questionnaire can contain up to 25 items. Twenty-three of the items can be selected from a catalogue or written by the instructor or department. Two are considered "global" and are hard-coded onto each ICES questionnaire as Item One and Two. They are the items considered for analysis in this study:

1. Rate the overall effectiveness of the instructor.
2. Rate the overall quality of the course.

The range of ratings is from 1 to 5 for both item 1 and item 2. Data from the University of Illinois at Urbana-Champaign campus ICES administration for two terms, fall 1994 and spring 1995, were used in this study. The total observations are 235,187. The instructors include faculty and teaching assistants. This study will be concerned with only U of I faculty.

Methods, Techniques or Modes of Inquiry

Course motivation

The ICES Instructor Report, produced for each set of completed evaluation forms, classifies overall course motivation as required, mixed, or elective. Each student selected his or her motivation for being in the course from three options. ICES labels the specifically required course as 3, required but a choice among several courses as 2, and the elective course as 1. Within each class, ICES calculates the mean of the motivation types to determine the nature of the course, required (3.00-2.35), mixed (2.34-1.69), or elective (1.68-1.00).

Course Level

Course level is indicated by three-digit numbers between one and four hundred, with the higher numbers reflecting courses designed for upper level students.

Class Size

Class size is the number of students in each class. Classes of fewer than five were deleted because the ratings from these class sizes tend to be unreliable ($r < .4$). According to the analysis of variance, it is better to set class size as a fixed variable rather than a random variable. So, class size needs to be broken into several levels.

In looking at the frequency distribution of class size, it is apparent that two-thirds (2538 out of 3754) of classes range between 5 and 25. The average number of classes in each class size from 5 to 20 students is 135. As to the class size above 30, the number of classes is less than 50 in each class size. For the size above 60, there are less than 10 classes in each size. When the class size is larger than 130 students, there are merely 0, 1, or 2 in each size.

In order to keep about the same number of classes for each level of class size, the factor class size was divided into 7 levels (5-8, 9-12, 13-16, 17-21, 22-29, 30-48, and above 49 students).

Discipline

Discipline includes eight Colleges: Agricultural, Consumer, and Environmental Sciences, Applied Life Studies, Commerce and Business Administration, Communications, Education, Engineering, Fine and Applied Arts, and Liberal Arts and Sciences. Because the College of Liberal Arts and Sciences is so large, it will contain several distinct subsets. For purposes of this study, University of Illinois codes

were used to separate these subsets into four groups. Discriminant analysis was used to check these four groups of the College of Liberal Arts and Sciences. The results indicate that these four groups are not suitable for this study. According to the average student ratings for these four groups, the group involved mathematics has lower ratings than other groups. Therefore, the group with mathematics will be treated as one isolated group and the rest of the groups set into one group. After rechecking these two groups by discriminant analysis, the result is acceptable.

In conclusion, there are three types of course motivation (required, mixed, elective), four course levels (100-400), seven types of class size, and 9 types of academic discipline (after the College of Liberal Arts and Sciences was split into two groups) constituting a total of 3754 courses, with 106,303 students.

ANOVA Procedure

In terms of the multi-dimensionality of student ratings (Marsh, 1984), the two global items were the dependent variables for separate ANOVA models. For the statistical analyses, two of each of the ANOVAs were derived to analyze student ratings: two 1-way random design ANOVAs, two 2-way to two 5-way unbalanced nested ANOVAs (Hicks, 1993; Kirk, 1982). For the first step, two separate 1-way ANOVAs were used for all classes to detect whether student ratings were different for each class. In the second procedure, course motivation was put into the model to create the unbalanced nested 2-way ANOVA, then one additional variable at a time was introduced into the model. For the 5-way unbalanced nested model (full model), the variables (or variable combinations) that were not significant were eliminated.

For a one-factor ANOVA, a way to deal with the varying numbers of classes within each class size level was derived by Drasgow (1983) at the University of Illinois.

Derivation of 2-way ANOVA

Generalizing from the 1-way ANOVA, we may describe the procedure for an unbalanced 2-way nested ANOVA, as follows:

2-way unbalanced nested model

$$Y_{ijk} = \mu + A_i + B_{j(i)} + \varepsilon_{k(ij)}$$

where μ = the overall mean
A_i = course motivations, i = 1, 2, 3
$B_{j(i)}$ = classes, j = 1,..., b_i for all i
$\varepsilon_{k(ij)}$ = random error, k = 1,..., n_{ij} for all i, j

A_i is fixed variable and $B_{j(i)}$ is random varible which is nested within A_i

Expected mean square (EMS) for equal sample size and balanced nested model:

$EMS_{within} = \sigma_\varepsilon^2$

$EMSB = \sigma_\varepsilon^2 + n\sigma_B^2$

$EMSA = \sigma_\varepsilon^2 + n\sigma_B^2 + bn\phi_A$

where $\phi_A = \frac{\sum_{i=1}^{3} A_i^2}{3-1}$

So, for the estimated variances

$\hat{\sigma}_\varepsilon^2 = MSE$

$\hat{\sigma}_B^2 = \frac{MSB - MSE}{n}$

$\hat{\phi}_A = \frac{MSA - MSB}{bn}$

Each class had different sample size, requiring the use of the concept of unequal sample size to do adjustment (Hicks, 1993). s′ is the adjusted sample size for each class:

$$s' = \left(\frac{1}{\sum_{i=1}^{3} b_i - 1}\right)\left(\sum_{i=1}^{3}\sum_{j=1}^{b_i} n_{ij} - \frac{\sum_{i=1}^{3}\sum_{j=1}^{b_i} n_{ij}^2}{\sum_{i=1}^{3}\sum_{j=1}^{b_i} n_{ij}}\right)$$

$$= \left[\frac{1}{\text{number of class} - 1}\right]\left[\text{total observations} - \left(\frac{\text{sum of squares of class sizes}}{\text{total observations}}\right)\right]$$

For each level of course motivations, there exist a different number of classes and each class has a different sample size. In order to get the general subtotal of observations for course motivations, doing adjustment is necessary. (b′s′) is the adjusted sample size for course motivation. Considering the concept of adjusted sample size for each class, the overall sample size may be adjusted by number of levels of course motivation to find (b′s′).

$$b's' = \left(\frac{1}{3-1}\right)\left(\sum_{i=1}^{3}\left(\sum_{j=1}^{b_i} n_{ij}\right) - \frac{\sum_{i=1}^{3}\left(\sum_{j=1}^{b_i} n_{ij}\right)^2}{\sum_{i=1}^{3}\left(\sum_{j=1}^{b_i} n_{ij}\right)}\right)$$

The estimated variances for unequal sample size and unbalanced nested model:

$$\hat{\sigma}_\varepsilon^2 = \text{MSE}$$

$$\hat{\sigma}_B^2 = \frac{\text{MSB} - \text{MSE}}{s'}$$

$$\hat{\phi}_A = \frac{\text{MSA} - \text{MSB}}{b's'}$$

The variance of class means:

$\sigma_{\bar{y}_{ij}}^2 = \text{Var}\left(\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} y_{ijk}\right)$ where n_{ij} is sample size for each class.

$$= \text{Var}\left(\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}\left(\mu + A_i + B_{j(i)} + \varepsilon_{k(ij)}\right)\right)$$

$$= \text{Var}\left(\mu + A_i + B_{j(i)} + \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}\varepsilon_{k(ij)}\right)$$

$$= \phi_A + \sigma_B^2 + \frac{\sigma_\varepsilon^2}{n_{ij}}$$

The estimated variance of class means:

$$\hat{\sigma}^2_{\bar{y}_{ij}} = \hat{\phi}_A + \hat{\sigma}^2_B + \frac{\hat{\sigma}^2_\varepsilon}{n_{ij}}$$

$$= \frac{MSA - MSB}{b's'} + \frac{MSB - MSE}{s'} + \frac{MSE}{n_{ij}}$$

The distribution for each class size is $N(\mu_{\bar{y}}, \phi_\alpha + \sigma^2_\beta + \frac{\sigma^2_\varepsilon}{n_{ij}})$

The test statistic is $z = \frac{\bar{y}_{ij} - \bar{\bar{y}}}{[\text{EstVar}(\bar{y}_{ij})^{1/2}]}$

Derivation of 5-way ANOVA

The mathematical rule was followed to extend the model to 5-way unbalanced nested ANOVA used in this research.

5-way unbalanced nested full model:

$$Y_{ijklm} = \mu + A_i + B_j + Z_k + D_l + AB_{ij} + AZ_{ik} + AD_{il} + BZ_{jk} + BD_{jl} + ZD_{kl} + ABZ_{ijk} + ABD_{ijl} + AZD_{ikl} + BZD_{jkl} + ABZD_{ijkl} + C_{m(ijkl)} + \varepsilon_{n(ijklm)}$$

where μ = the overall mean

A_i = course motivations, i = 1, 2, 3

B_j = course levels, j = 1,..., 4

Z_k = class sizes, k = 1,..., 7

D_l = disciplines, l = 1,..., 11

$C_{m(ijkl)}$ = classes, m = 1,..., c_{ijkl} for all i, j, k, l

$\varepsilon_{n(ijklm)}$ = random error, n = 1,..., n_{ijklm} for all i, j, k, l, m

A_i, B_j, Z_k and D_l are fixed variables and $C_{m(ijkl)}$ is random varible which is nested within A_i, B_j, Z_k and D_l.

Summary

The factors of course motivation, course level, class size and discipline are fixed variables. For the analyses of variance, the nested factor, individual class, is a random variable.

The ANOVA model is based on a normal distribution. Thus, the ratings are assumed to be normally distributed, with the mean and deviation given above.

The class size is the total number of students in each class. In accord with the central limit theorem, the standard error should be divided by each class size.

According to the results of the unbalanced nested ANOVA, the mean and adjusted standard deviation were computed for the distribution. Each level of class size has a separate distribution. The mean differences were compared within the two variables, course motivation and class size.

Results

All of my statistical analyses of results from ICES are separated into item 1 (course effectiveness) and item 2 (course quality). Generally, student ratings of teaching effectiveness are higher than course quality. That is, most students have higher satisfaction with teaching effectiveness than with course quality.

The descriptive statistics for course motivation, course level, class size, and discipline are reported. Among each level of every variable, number of classes, average of student ratings, standard deviation, minimum score, and maximum score are compared.

The analyses of variance for teaching effectiveness and course quality are presented. As a result of using ANOVA, the intraclass correlation is interpreted for both items. Whether or not the ANOVA model for teaching effectiveness and course quality is consistent is shown. In order to indicate the relationship of variables with a higher order interaction effect, some figures and descriptive statistics are presented.

A strong association between the dependent variable and the independent variable indicates a significant effect. The strength of association for significant treatment effects is estimated and in accord with this estimation, the proportion of variance in the assessment of teaching effectiveness or course quality is provided.

The distribution of class size is arranged by course motivation for item 1 and item 2. The descriptive statistics for course motivation and class size are appended to the distribution tables. There are four percentage categories (low 10%, middle 30% and 70%, and high 90%). Generally, the larger the class size, the lower the rating and

the more narrow the range of ratings. Also, the differences between the distribution of teaching effectiveness and course quality are compared.

Descriptive Statistics

Course motivation

There are three types of course motivation: required, mixed, and elective. According to Table 1, the average ratings for elective courses (4.25 for item 1 and 4.23 for item 2) are higher than those for mixed courses (4.13 for item 1 and 4.08 for item 2); the average ratings of mixed courses are higher than those of required courses (4.01 for item 1 and 3.93 for item 2). As to the number of classes for each category, there are more required courses than either elective or mixed courses.

Course Level

As seen in Table 2, there are more 300-level courses than courses at other levels. For both teaching effectiveness and course quality, the higher the level of the course, the higher the ratings average. That is, there is a positive correlation between course level and student ratings.

Class Size

There are seven categories for class size: 5-8 students, 9-12, 13-16, 17-21, 22-29, 30-48, and 49 and above. In general, there is a negative correlation between class size and average student ratings; the smaller the size of the class, the higher the ratings (see Table 3). However, maximum rating rises slightly in the very large class category (49 students or more), after reaching a low in the large class category (30-48 students). Thus, while mean rating declines steadily as class size increases, maximum

rating becomes a U-shaped curve. For some reason, one or more very large classes have higher ratings than the large classes.

Discipline

There are nine categories for discipline including Agricultural, Consumer, and Environmental Sciences, Applied Life Studies, Commerce and Business Administration, Communications, Education, Engineering, Fine and Applied Arts, and Liberal Arts and Sciences (AREA I and AREA II). According to the average ratings for each discipline (Table 4), the College of Fine and Applied Arts has a higher rating than other colleges. The Colleges of Agricultural, Consumer, and Environmental Sciences and Education have the next highest ratings; the Colleges of Engineering and Liberal Arts and Sciences (AREA II: related to mathematics) have lower ratings than other colleges. With the exception of the College of Liberal Arts and Sciences, the College of Engineering has the greatest number of courses at the University of Illinois.

Analysis of Variance

All treatment effects are significant. The 4-way interaction terms are insignificant (p-value>0.05) for both teaching effectiveness and course quality. ANOVA results for item 1 and 2 are summarized in Table 5 and 6. Generally, the results of test statistics for teaching effectiveness and course quality are consistent. As to the 3-way interaction terms, only the term composed of motivation, course level, and discipline is strongly significant (p-value<0.0001). Therefore, course motivation, course level, and discipline will affect each other. Thus, descriptive statistics for motivation, level and discipline are needed.

The results of 2-way interaction terms show a correlation between the elements of three pairs of variables: course motivation and discipline, course level and class size, and course level and discipline. There is an especially strong interaction between course level and discipline. There is no significant interaction between the elements of three other pairs: course motivation and course level, course motivation and class size, and class size and discipline. Because the class size factor is not in the 3-way significant interaction term, descriptive statistics for course level plus class size are presented. The four main factors course motivation, course level, class size, and discipline are significant. The differences among the individual classes are also strongly significant.

The R-squared value, the proportion of the model explanation, is 39% for effectiveness of instructors and 31% for quality of course. That means the ANOVA model can explain 39% of the variance of student ratings for teaching effectiveness and 31% of the variance of student ratings for course quality. In conclusion, differences in course motivation, course level, class size, and discipline will affect the results of student ratings.

The Strength of Association for Variance

Distinctions between individual classes account for 31.88% of variance of student ratings for teaching effectiveness and 21.97% of variance of student ratings for course quality. Therefore, class differences are the major factor for both teaching effectiveness and course quality.

Based on course difference (Table 7), course motivation accounts for 2.85% of variance for teaching effectiveness and 5.34% of variance for course quality.

Discipline accounts for 3.48% of variance for teaching effectiveness and 6.24% of variance for course quality. Class size accounts for 3.26% of variance for teaching effectiveness and 4.55% of variance for course quality. Thus, compared with course motivation and discipline, class size is the second most important factor for teaching effectiveness and the third most important factor for course quality. Other than course motivation, class size and discipline, the associated strengths of variance for the rest of the main factors and significant interaction terms are less than 2%.

Descriptive Statistics for 3-way interaction term (motivation, level, and discipline) (Figure 1-9)

Generally, with different combinations of course motivation, level, and discipline, the rating patterns will be different except in 300-level courses. Based on 300-level courses, the average student rating for elective courses is higher than that for mixed courses; the average student rating for mixed courses is higher than that for required courses.

At the College of Communications, there are not very many classes in any of combinations of course motivation and level except in 300-level required courses (44; about 54% of total classes). In addition, the College of Applied Life Studies does not have many classes in mixed and elective courses. Therefore, due to the limited information from the small sample size, the differences for combinations of course motivation and level of the College of Communications and the mixed and elective courses of the College of Applied Life Studies will not be compared.

For the College of Agricultural, Consumer, and Environmental Sciences, average ratings for 400-level required courses are higher than those for other required

courses. As to mixed courses of this college, the average ratings for 100- and 200-level courses are higher than those for 300- and 400-level courses. There is no large difference among the levels of elective courses in the College of Agricultural, Consumer, and Environmental Sciences. For 100-level courses, the average rating of required courses is the lowest for course quality, but not much lower than that of the other types of motivation in the area of teaching effectiveness.

In the College of Education all of the course levels follow the same rating pattern: the average rating of required courses is lower than that of mixed courses, and the average rating of mixed courses is lower than that of elective courses.

The College of Liberal Arts and Sciences (AREA I) has the same pattern, but there is no large rating difference due to course motivation for 200 level. For the College of Agricultural, Consumer, and Environmental Sciences and College of Commerce and Business Administration, except in 400 level, all of course levels follow the same pattern. The College of Engineering follows the same pattern except in 200-level courses.

Eighty-four percent of courses in the College of Commerce and Business Administration are 300 and 400 level. There are not many 100- and 200-level mixed and elective courses. For this college, the rating of 200-level required courses is the lowest. There are no large rating differences among course motivations for 300-level courses. As to 400-level courses, the rating of mixed courses is the highest.

For the College of Education, there are few 100- and 200-level courses (<6 in each category) except for 200-level required courses (20). Generally, the student ratings are high.

As to the College of Engineering, there are a large number of lower level required courses (120 classes at the 100 level and 146 classes at the 200 level) and higher level elective courses (156 classes at the 300 level and 109 classes at the 400 level). Generally, the ratings of required courses are lower and the ratings of elective courses are higher. For required courses in the College of Engineering, the higher the course level, the higher the student ratings, but mixed and elective courses do not follow this rating pattern.

Generally, the average of the student ratings in the College of Fine and Applied Arts is higher than in other disciplines because the average ratings for all of the combinations of motivation and level are higher than 4.00. There is a strong difference between 100-level mixed classes and 400-level mixed classes (t=2.42). However, there are no large rating differences between required and elective courses. There are no differences within course motivation for the 200 to 400 levels.

As to the College of Liberal Arts and Sciences (AREA I), there are no large differences within the level of courses for mixed courses. The highest rating is in 400-level elective courses, the lowest rating is in 100-level required courses.

Generally, the average ratings in the Liberal Arts and Sciences (AREA II) were low, especially for required and mixed courses. Most of the courses are required courses (56%). On the other hand, there are not many courses among lower level elective courses and 200- and 400-level mixed courses. There are no strong differences within the levels of courses for required courses and those for mixed courses regarding the rating of teaching effectiveness. For 300-level courses, there are

no large ratings differences for required and mixed courses. The average rating of 300-level elective courses is higher than both 300-level required and mixed courses.

Descriptive Statistics for Course Level and Class Size

Because the factors of course level and class size are highly related, different combinations of course motivation and class size will have different rating patterns. In particular, very large 400-level classes (49 and more) have a very low average rating, but the sample size is a little small (15). On the other hand, the ratings of 200-level courses are higher than those of 100-level courses in each of the class size categories.

Generally, at all of the course levels, the classes with 5-8 students have the highest average ratings for course quality except at the 300, where classes with 9-12 students have slightly higher average ratings than other larger or smaller class sizes (as shown in Table 8). The largest classes (49 students or more) have the lowest ratings except at the 100 level, where classes with 30–48 students have the lowest average ratings for teaching effectiveness and course quality.

In the ratings of teaching effectiveness and course quality, at the 100 level, there is a U-shaped trend for the factor of class size (see Figure10).

Distribution

The table of the distribution of each size category shows that the ratings of required courses are lower than those of mixed courses, and the ratings of mixed courses are lower than those of elective courses (Table 9). This verifies that course motivation and class size are not highly related. For each course motivation, there is a linear trend among the class sizes except in the elective courses, where there is a U-shaped trend.

Because of the uneven sample sizes for each category, the standard deviation, and class size of each category and common standard deviation (produced by 5-way unbalanced nested ANOVA) are included to develop the adjusted standard deviation. The different combinations of course motivation and class size have different distributions. Every distribution has a different mean (average of student ratings) and standard deviation. Therefore, the estimated ratings and range of the ratings would be expected to differ by class size. Usually, the small class has a wider range and the very large class has a more narrow range.

Summary

Course motivation, course level, class size and discipline affect student ratings for both teaching effectiveness and course quality. Among these four factors, course motivation, class size and discipline play the most important role. Also, discipline has an effect on course motivation. Generally, ratings in elective courses are higher than those in mixed courses; ratings in mixed courses are higher than those in required courses for most of the colleges except the College of Communications and the College of Liberal Arts and Sciences. As to the College of Communications and the College of Liberal Arts and Sciences, required courses have higher ratings than mixed courses.

There is a notable interaction among course motivation, course level, and discipline. In different combinations of the motivations, levels, and disciplines, the patterns of student ratings will be different, except in 300-level courses. Class size does affect student ratings. The only factor that will alter the effect of class size is course level. Usually, for each course level, the very small classes have the highest

ratings; the very large classes have the lowest ratings except in 100-level courses where the lowest ratings are in the large classes (30-48).

For the distribution of class size and course motivation, the larger the class size, the lower the ratings and the more narrow the range of ratings. There is no interaction term between course motivation and class size because ratings in elective courses are the highest and ratings in required courses are the lowest for all sizes of classes. Moreover, there is a nonlinear relationship among class size in elective courses, while, there is a linear relationship among class sizes for required and mixed courses.

In conclusion, there are many factors that will affect student ratings. Individual differences among the classes are the main factor. Besides this difference of individual class, discipline has the next most powerful effects. Course motivation and class size have some influence on student ratings, but not as powerful as discipline.

Discussion

Student ratings of instructors and courses will continue to be important aspects of the faculty evaluation process. Studies such as this one provide insights that can be used to improve the instructional setting, course design, and course delivery.

Course motivation does have an impact on student ratings. The question is whether class size plays an important role in affecting student ratings. Ratings of instructors who teach in the small classes tend to be higher than those of instructors who teach in the very large class. Accounting for these systematic differences if they are sizable is critical to the measurement of teaching effectiveness.

This study uses the evaluation questionnaire from the University of Illinois, the Instructor and Course Evaluation System (ICES). The range of student ratings is from 1 to 5 and the range in class size is from 1 to 726. In different colleges and different courses, the form of the ICES questionnaires varies, but there are two common items: teaching effectiveness for Item One and course quality for Item Two. These two common items were the focus of this study of student ratings. Overall, students are more satisfied with teaching effectiveness than with course quality.

The unbalanced nested ANOVA has been applied to extend our understanding of influences on teaching effectiveness and course quality assessed with ICES. There are five factors included: course motivation (required, mixed, and elective), course level (100-400 level), class size (5-8, 9-12, 13-16, 17-21, 22-29, 30-48, and 49 and more), academic discipline (8 colleges are divided into 9 categories), and individual class differences. The relationships between the 5 factors and student ratings of teaching effectiveness and course quality are studied according to the ANOVA result.

The current ICES reporting format accommodates course motivation only. In this section, the joint distributions of class size and course motivation are developed and compared to the current ICES report. The implication of the differences between the new report and old report are presented.

Summary Result from 5-way Unbalanced Nested ANOVA

The results of the ANOVA show that course motivation, course level, class size, discipline, and individual class do have an impact on student ratings of both teaching effectiveness and course quality. There is an interaction term for course motivation, class level and discipline. Thus, different combinations of course motivation, level, and discipline display different patterns of student ratings. The significant 2-way interactions are course motivation and discipline, course level and class size, and course level and discipline. That is, in each of these pairs, one of the factors alters the effect of the other. Course level is the only factor that interacts with class size.

Based on the strength of association for variance, individual differences among courses are the major effect on student ratings. Moreover, course motivation, class size and discipline have secondary effects. In terms of the large sample size (106,303), the rest of the factors and significant interaction terms only account for less than 2% of variance in student ratings.

Summary Results from Descriptive Statistics for Factors

Generally, the ratings of elective courses are higher than those of mixed courses; the ratings of mixed courses are higher than those of required courses. As to the factor of course level, the higher level classes have higher ratings than the lower

level classes. Student ratings in small classes are higher than those in large classes. By academic discipline, ratings in the fine and applied arts courses higher than those in the engineering and math-related courses.

Actually, the rating pattern will be different in different combinations of course motivation, class level, and discipline because these factors are highly interacted. In addition, the 300-level courses keep the same rating pattern (the average rating of elective courses is the highest; the average rating of required courses is the lowest) of course motivation for all disciplines.

There is an interaction term between course level and class size. In different combinations of course level and class size the rating patterns will be different. In general, small classes (5-8) have the highest ratings for both teaching effectiveness and course quality. As to the lowest rating, the very largest classes (49 and more) have the lowest ratings except for the 100-level courses. Therefore, a U-shaped trend occurs in 100-level courses.

For the overall combinations of course level and class size, the average rating of very large 400-level classes is very low. Thus, the rating range at the 400 level classes is wider than the range at other levels.

Implication of the Distribution

The current ICES faculty report considers course motivation only. The distribution of this study, however, has both course motivation and class size because class size does affect student ratings. In the table 9 showing the distribution a faculty member can easily find his percentage score of student ratings.

The adjusted standard deviation for each distribution is based on the class size and standard deviation from the combined categories of course motivation and class size and on the standard deviation from the 5-way unbalanced nested ANOVA. According to the different average of student ratings, class size, and standard deviation for each combined category of course motivation and class size, every distribution will have a different mean and standard deviation.

The advantages of these distributions are shown below.

(1) The table of distributions with percentiles provides more information than the current ICES faculty report.

(2) Based on average ratings, class size, and deviation from the combination of class size and course motivation and deviation from the 5-way unbalanced nested ANOVA, the distributions can present the rating pattern more accurately than the current ICES faculty report.

There is no interaction term between course motivation and class size. Therefore, the smaller the class size, the higher the ratings except in elective courses where there is a U-shaped trend for class size.

Limitations and Problems

The range of student ratings is 1 to 5. Ninety-five percent of the classes have average student ratings equal to or more than 3. The range of class size of effective data is 5 to 726. Seventy-five percent of classes have 30 students or less in each class. The frequency distribution of class size is like the Poisson distribution or normal distribution with positive skewness. Therefore, the range of student ratings is very narrow. The range of class size is very wide, but most classes are not large classes.

The ratings among the percentage categories are very close. In other words, there are lots of courses with the same ratings.

Some colleges do not have many classes in some combinations of course motivation and class size, specially the College of Communications and the College of Applied Life Studies. Due to small sample size, comparing the rating patterns in these two colleges would not be reliable, nor would it be worthwhile to generate the distribution of class size for a 3-way significant interaction term (course motivation, class level, and discipline).

In addition, this study utilized ICES fall 1994 and spring 1995 data. It could be helpful to use more than one academic year's evaluation data because the number of classes for each discipline will be greater. The limitation of the small samples from the small colleges might be eliminated.

Recommendations for Future Research

The student rating scale is a 5-point scale. Most classes (95%) are higher than 3 on average ratings. There are about 3,570 classes that have average ratings between 3 and 5. Thus, there are little rating differences among different classes. Due to insufficient point spread on the rating scale, much important information may be lost and the accuracy of the results may be reduced. Adding more numbers to the point scale is likely to increase the reliability and accuracy of the analyses of variance (Braskamp & Ory, 1994).

The R-squared values of the 5-way unbalanced nested ANOVA are a little low (39% for teaching effectiveness and 31% for course quality). It is possible that the spread in the point scale is too small or some important factors are excluded from the

ANOVA models. For example, some researchers found a few important factors that may influence student ratings, including gender of instructor (Basow & Silberg, 1987; Bennett, 1982; Kierstead, D'Agostino, & Dill, 1988; Marsh, 1987), academic achievement or research productivity (Aleamoni & Yimer, 1973; Feldman, 1987; Marsh, 1984; Simpson, 1995), personality of instructor (Feldman, 1986; Marsh, 1987; Murray, Rushton, & Paunonen, 1990), expected grade (Marsh, 1984, 1987) and workload or class preparing time (Franklin & Theall, 1995; Gillmore, 1994; Marsh, 1984, 1987). Wigington, Tollefson, and Rodriguez (1989) also found that professors received lower ratings than assistant and associate professors. Therefore, adding more factors into the ANOVA model may let the ANOVA results be more reliable and provide more information.

References

Aleamoni, L. M., & Graham, M. H. (1974). The relationship between CEQ ratings and instructor's rank, class size and course level. Journal of Educational Measurement, 11, 189-201.

Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? Journal of Educational Psychology, 79, 308-314.

Bausell, R. B., & Bausell, C. R. (1979). Student ratings and various instructional variables from a within-instructor perspective. Research in Higher Education, 11, 167-177.

Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: evidence relating to the question of gender bias in teaching evaluation. Journal of Education Psychology, 74, 170-179.

Benz, C. R., & Blatt, S. J. (1996). Meanings underlying student ratings of faculty. The Review of Higher Education, 19, 411-433.

Braskamp, L. A., & Ory, J. C. (1994). Assessing faculty work. San Fransisco: Jossey-Bass.

Cashin, W. E., & Downey, R. G. (1992). Using global student rating items for summative evaluation. Journal of Education Psychology, 84, 563-572.

Cashin, W. E., & Downey, R. G. (1995). Disciplinary differences in what is taught and in students' perceptions of what they learn and of how they are taught New Directions for Teaching and Learning, no 64, 81-92.

Drasgrow, F. (1983). Class Draft. University of Illinois at Urbana-Champaign.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. Research in Higher Education, 9, 199-242.

Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. Research in Higher Education, 10, 149-172.

Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. Research in Higher Education, 24, 139-213.

Feldman, K. A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: a review and exploration. Research in Higher Education, 26, 227-298.

Franklin, J., & Theall, M. (1995). The relationship of disciplinary differences and the value of class preparation time to student ratings of teaching. New Directions for Teaching and Learning, no 64, 41-48.

Haladyna, T., & Hess, R. K. (1994). The detection and correction of bias in student ratings of instruction. Research in Higher Education, 35, 669-687.

Hicks, C. R. (1993). Fundamental concepts in the design of experiments. New York: Saunders college publishing.

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. Journal of Educational Psychology, 77, 187-196.

Kierstead, D., D'Agostino, P., & Dill, H. (1988). Sex role stereotyping of college professors: Bias in students' ratings of instructors. Journal of Educational Psychology, 80, 342-344.

Kirk, R. E. (1982). Experimental design. Brooks/Cole Publishing Company.

Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. Review of Research in Education, 3, 211-240.

Marsh, H. W. (1984). Students' evaluation of university teaching: Dimensionality, reliability, validity, potential biases, and utility. Journal of Educational Psychology, 1984, 76, 707-754.

Marsh, H. W. (1987). Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388.

Murray, H. G., & Renaud, R. D. (1995). Disciplinary differences in classroom teaching behaviors. New Directions for Teaching and Learning, no 64, 31-39.

Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. Journal of Educational Psychology, 82, 250-261.

Newport, J. F. (1996). Rating teaching in the USA: Probing the qualifications of student raters and novice teachers. Assessment & Evaluation in Higher Education, 21, 17-21.

Simpson, R. D. (1995). Uses and misuses of student evaluations of teaching effectiveness. Innovative Higher Education, 20, 3-5.

Wigington, H., Tollefson, N., & Rodriguez, E. (1989). Students' ratings of instructors revisited: Interactions among class and instructor variables. Research in Higher Education, 30, 331-344.

Table 1

Descriptive Statistics for Course Motivation

Motivation	N	Mean	Std Dev	Min	Max
Teaching Effectiveness					
Required	1572	4.01	0.63	1.58	5.00
Mixed	1097	4.13	0.55	1.92	5.00
Elective	1085	4.25	0.51	2.00	5.00
Course Quality					
Required	1574	3.93	0.58	1.91	5.00
Mixed	1096	4.08	0.51	2.03	5.00
Elective	1088	4.23	0.49	2.30	5.00

Table 2

Descriptive Statistics for Course Level

Course Level	N	Mean	Std Dev	Min	Max
Teaching Effectiveness					
100 level	763	4.03	0.63	1.58	5.00
200 level	853	4.10	0.59	1.94	5.00
300 level	1351	4.10	0.57	1.70	5.00
400 level	787	4.26	0.53	2.17	5.00
Course Quality					
100 level	763	3.94	0.57	1.92	5.00
200 level	853	4.03	0.55	2.03	5.00
300 level	1352	4.05	0.53	1.91	5.00
400 level	790	4.23	0.51	2.17	5.00

Table 3

Descriptive Statistics for Class Size

Class Size	N	Mean	Std Dev	Min	Max
Teaching Effectiveness					
5-8	573	4.33	0.49	2.29	5.00
9-12	610	4.26	0.54	1.58	5.00
13-16	539	4.15	0.56	2.25	5.00
17-21	530	4.09	0.58	1.70	5.00
22-29	521	4.05	0.57	2.11	5.00
30-48	530	3.97	0.61	1.92	4.96
49 and up	451	3.90	0.63	1.79	4.99
Course Quality					
5-8	577	4.30	0.50	2.57	5.00
9-12	619	4.23	0.50	1.92	5.00
13-16	529	4.11	0.54	2.13	5.00
17-21	533	4.01	0.55	1.95	5.00
22-29	517	3.98	0.53	2.17	4.97
30-48	531	3.90	0.53	2.30	4.89
49 and up	452	3.82	0.52	1.91	4.95

Table 4

Descriptive Statistics for Discipline

College	N	Mean	Std Dev	Min	Max
Teaching Effectiveness					
ACES	363	4.24	0.51	2.25	5.00
ALS	153	4.16	0.57	1.92	5.00
CBA	419	4.01	0.59	1.97	5.00
COM	82	4.12	0.54	1.79	4.92
EDU	230	4.24	0.56	2.06	5.00
ENG	807	3.96	0.60	1.94	5.00
FAA	493	4.31	0.56	1.71	5.00
LAS1	906	4.16	0.56	1.70	5.00
LAS2	301	3.97	0.61	1.58	5.00
Course Quality					
ACES	365	4.23	0.46	2.43	5.00
ALS	153	4.07	0.54	2.42	5.00
CBA	419	3.97	0.55	2.09	5.00
COM	82	4.08	0.49	1.91	4.92
EDU	230	4.19	0.55	2.22	5.00
ENG	808	3.87	0.56	2.03	5.00
FAA	494	4.30	0.50	1.95	5.00
LAS1	906	4.10	0.51	2.00	5.00
LAS2	301	3.89	0.54	1.92	5.00

Table 5

Full Model (Dependent variable = Teaching effectiveness)

Source	DF	Type I SS	Mean Square	F-Value
CM	2	1436.102	718.051	70.442 **
Level	3	420.664	140.221	13.756 **
Size	6	883.709	147.285	14.449 **
Discipline	8	1389.159	173.645	17.035 **
CM*Level	6	18.625	3.104	0.305
CM*Size	12	196.676	16.390	1.608
CM*Displn	16	302.887	18.930	1.857 *
Level*Size	18	395.052	21.947	2.153 **
Level*Displn	24	778.412	32.434	3.182 **
Size*Displn	48	473.677	9.868	0.968
CM*Level*Size	34	279.119	8.209	0.805
CM*Level*Displn	44	1019.992	23.182	2.274 **
CM*Size*Displn	94	1040.812	11.072	1.086
Level*Size*Displn	119	1103.104	9.270	0.909
CM*Level*Size*Displn	109	664.693	6.098	0.598
Class(CM Level Size Displn)	3210	32721.233	10.194	15.447 **
Student(Class)	102549	67542.664	0.659	

$R^2 = 0.3897$

Table 6

Full Model (Dependent variable = Course quality)

Source	DF	Type I SS	Mean Square	F-Value
CM	2	2217.873	1108.937	162.510**
Level	3	605.236	201.745	29.565**
Size	6	976.537	162.756	23.851**
Discipline	8	1991.660	248.958	36.484**
CM*Level	6	24.628	4.105	0.602
CM*Size	12	110.731	9.228	1.352
CM*Displn	16	316.816	19.801	2.902**
Level*Size	18	280.914	15.606	2.287**
Level*Displn	24	804.205	33.509	4.911**
Size*Displn	48	332.079	6.918	1.014
CM*Level*Size	34	256.924	7.557	1.107
CM*Level*Displn	44	834.558	18.967	2.780**
CM*Size*Displn	94	754.574	8.027	1.176
Level*Size*Displn	119	849.794	7.141	1.047
CM*Level*Size*Displn	109	488.165	4.479	0.656
Class(CM Level Size Displn)	3214	21931.647	6.824	9.747**
Student(Class)	102408	71695.961	0.700	

$R^2 = 0.3137$

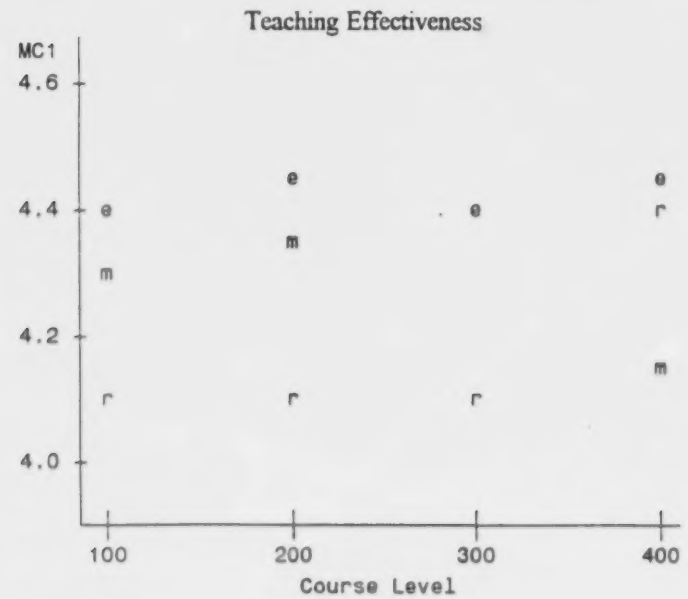
Table 7

Strength of Association for ICES Item 1 and Item 2 Presenting Course Difference

	Effect (%)	
Variable	Item 1	Item 2
1-way		
CM	2.85	5.34
Level	1.24	1.98
Size	3.26	4.55
Discipline	3.48	6.24
2-way		
CM*Level	----	----
CM*Size	----	----
CM*Discipline	0.00	0.30
Level*Size	0.69	0.61
Level*Discipline	0.82	1.66
Size*Discipline	----	----
3-way		
CM*Level*Size	----	----
CM*Level*Discipline	1.25	1.67
CM*Size*Discipline	----	----
Level*Size*Discipline	----	----

Note: Item 1 is teaching effectiveness. Item 2 is course quality.

College of Agricultural, Consumer, and Environmental Sciences



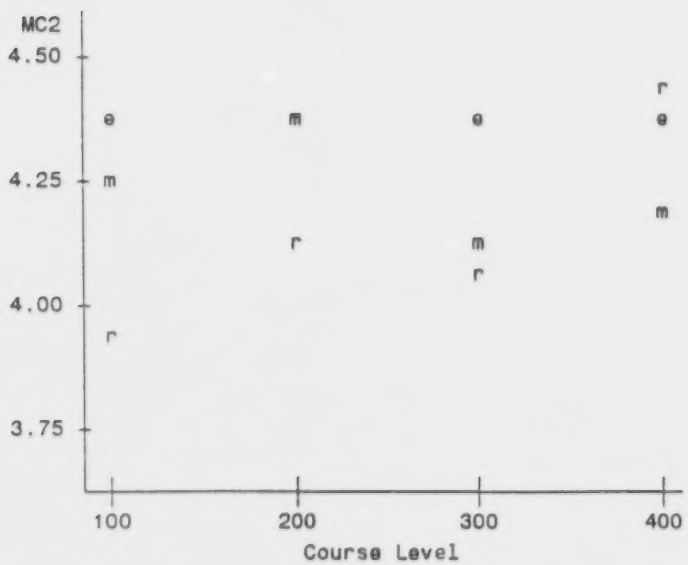
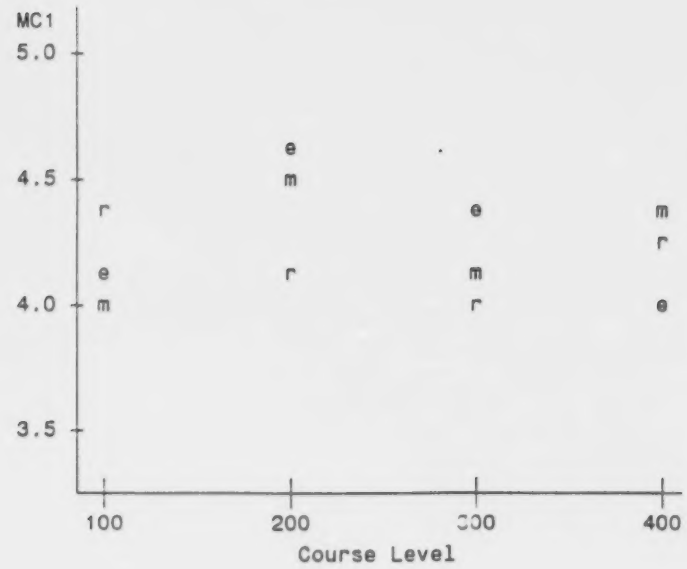


Figure 1. Scatter plots of the descriptive statistics for course motivation, course level and College of Agricultural, Consumer, and Environmental Sciences. Symbol is the value of course motivation. The mc1 and mc2 are the average of the class ratings.

College of Applied Life Studies

Teaching Effectiveness



Course Quality

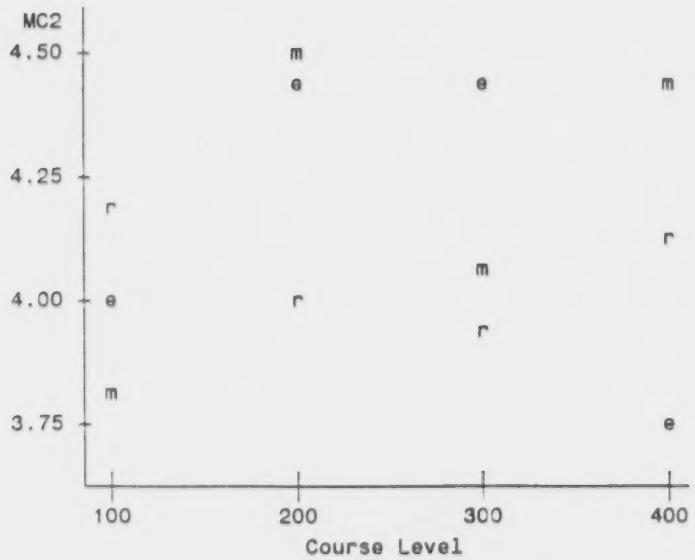
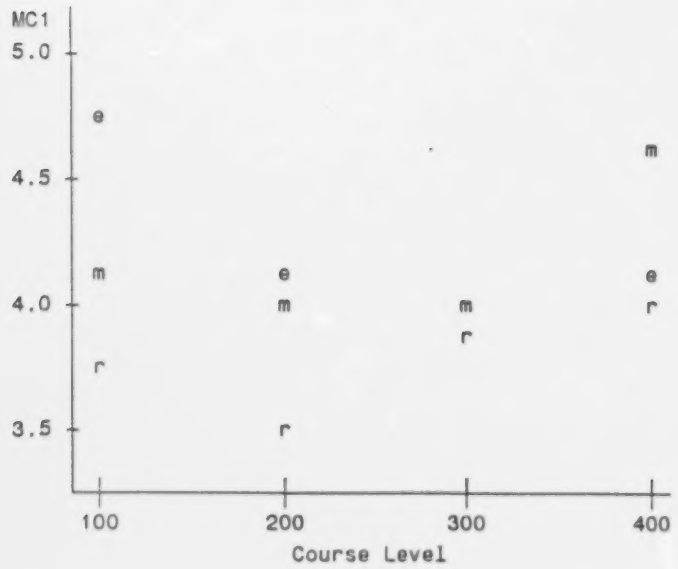


Figure 2. Scatter plots of the descriptive statistics for course motivation, course level and College of Applied Life Studies. Symbol is the value of course motivation.

College of Commerce and Business Administration

Teaching Effectiveness



Course Quality

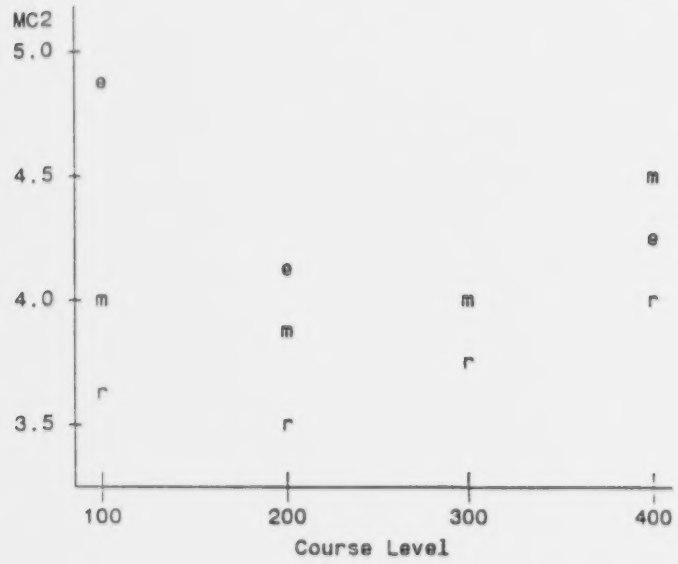
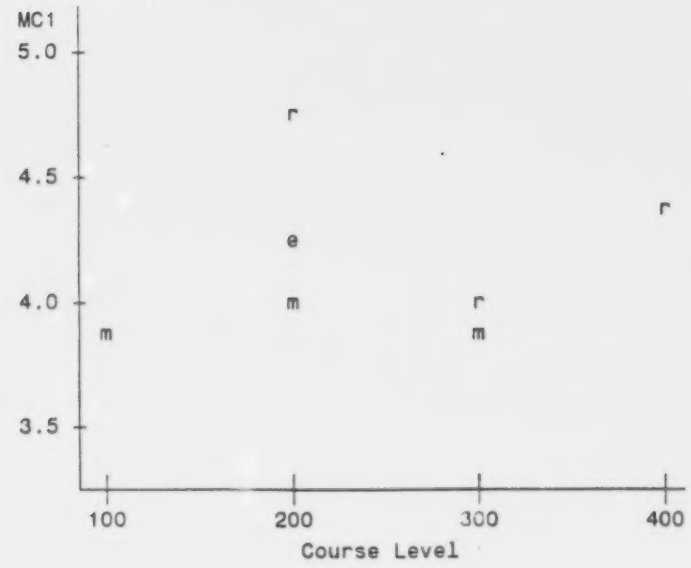


Figure 3. Scatter plots of the descriptive statistics for course motivation, course level and College of Commerce and Business Administration. Symbol is the value of course motivation.

College of Communications

Teaching Effectiveness



Course Quality

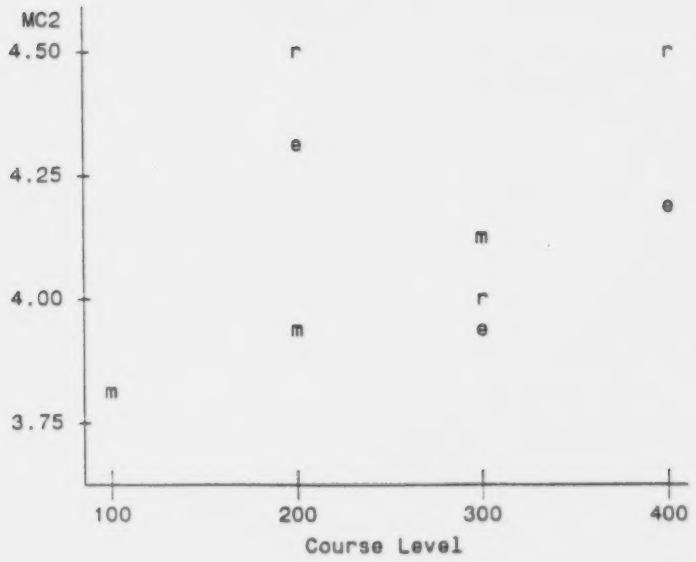
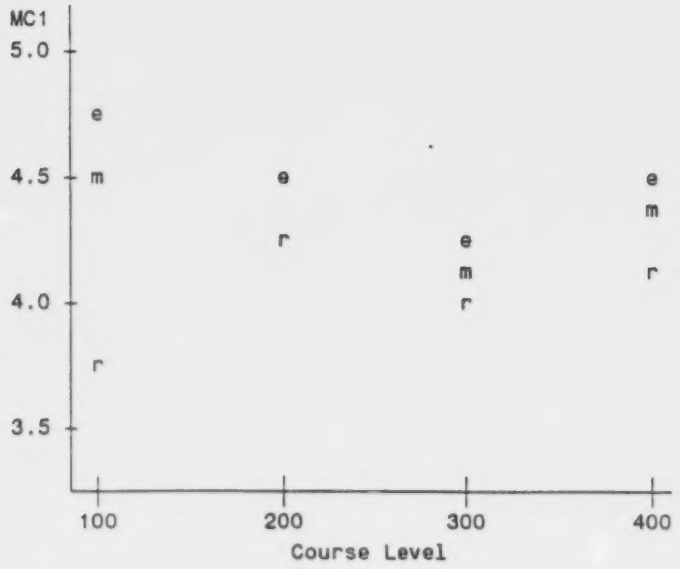


Figure 4. Scatter plots of the descriptive statistics for course motivation, course level and College of Communication. Symbol is the value of course motivation.

College of Education

Teaching Effectiveness



Course Quality

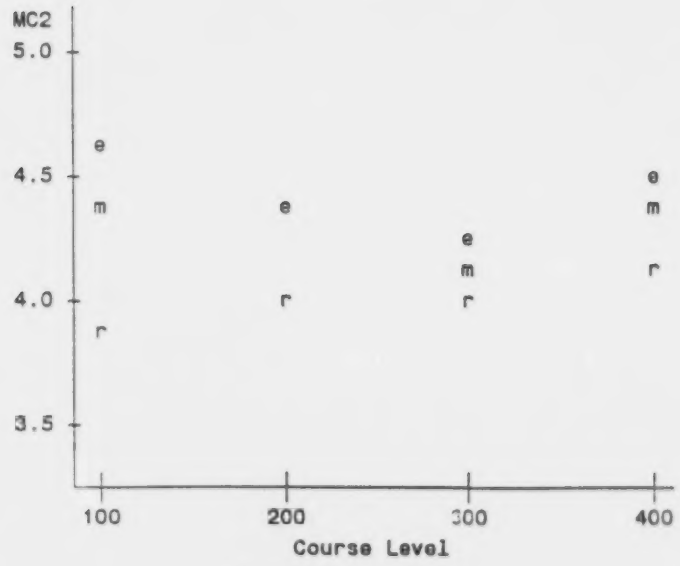
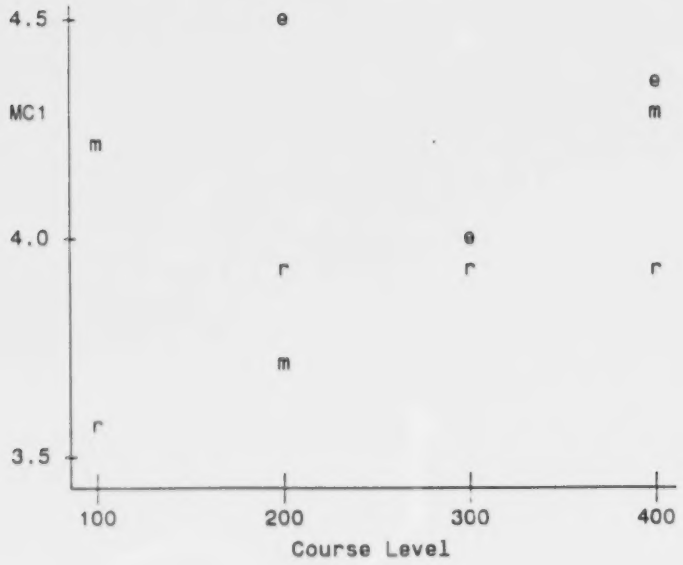


Figure 5. Scatter plots of the descriptive statistics for course motivation, course level and College of Education. Symbol is the value of course motivation.

College of Engineering

Teaching Effectiveness



Course Quality

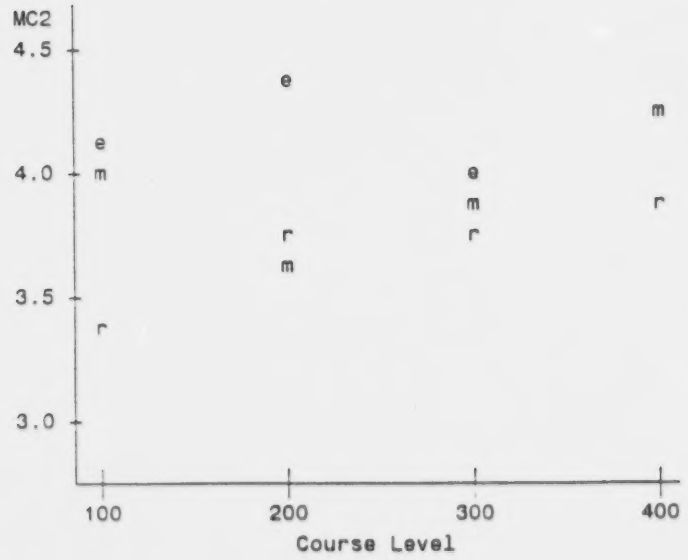
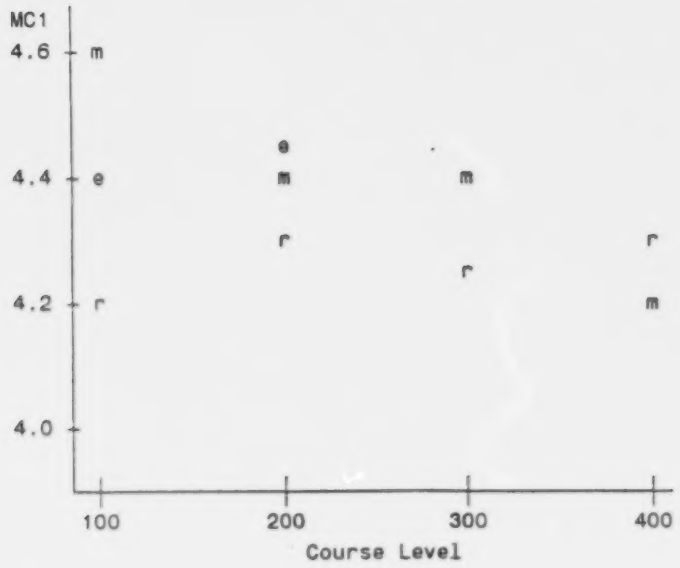


Figure 6. Scatter plots of the descriptive statistics for course motivation, course level and College of Engineering. Symbol is the value of course motivation.

College of Fine and Applied Arts

Teaching Effectiveness



Course Qua'ity

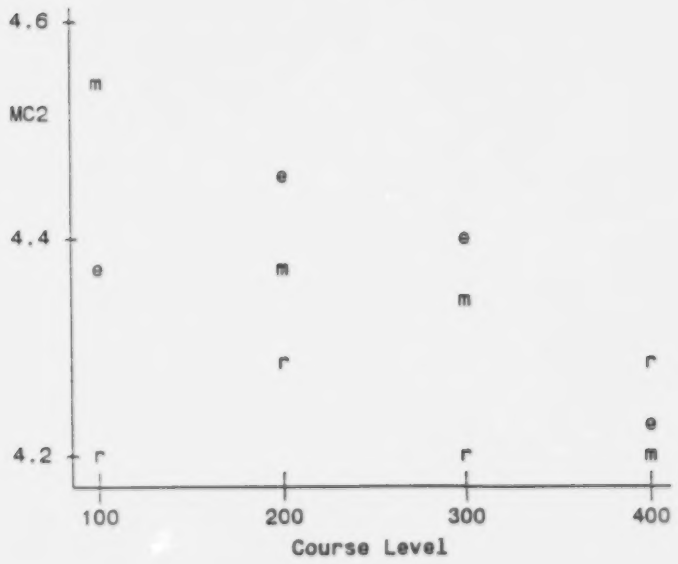
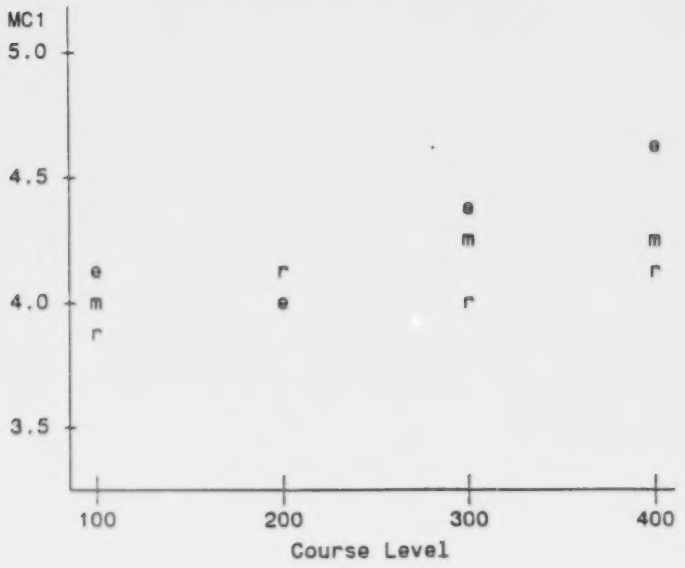


Figure 7. Scatter plots of the descriptive statistics for course motivation, course level and College of Fine and Applied Arts. Symbol is the value of course motivation.

College of Liberal Arts and Sciences I

Teaching Effectiveness



Course Quality

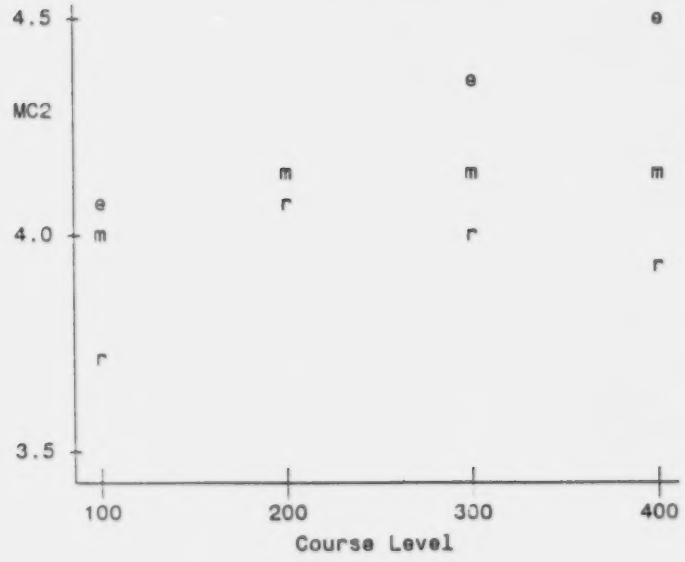
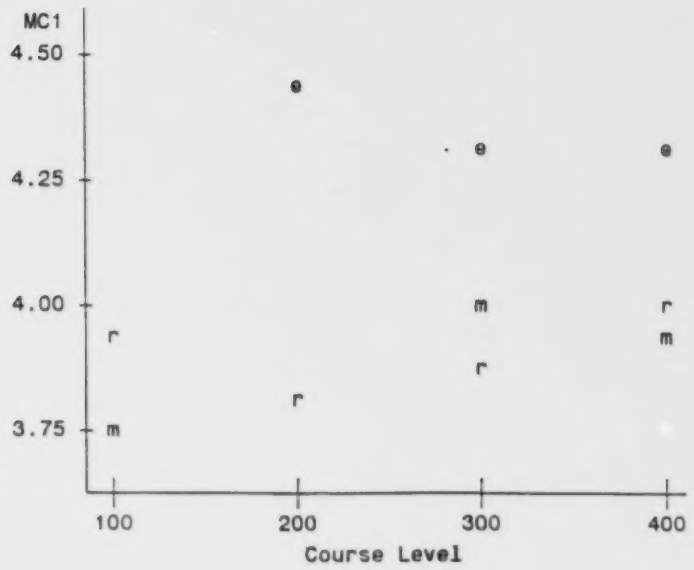


Figure 8. Scatter plots of the descriptive statistics for course motivation, course level and College of Liberal Arts and Science: AREA I. Symbol is the value of course motivation.

College of Liberal Arts and Sciences II

Teaching Effectiveness



Course Quality

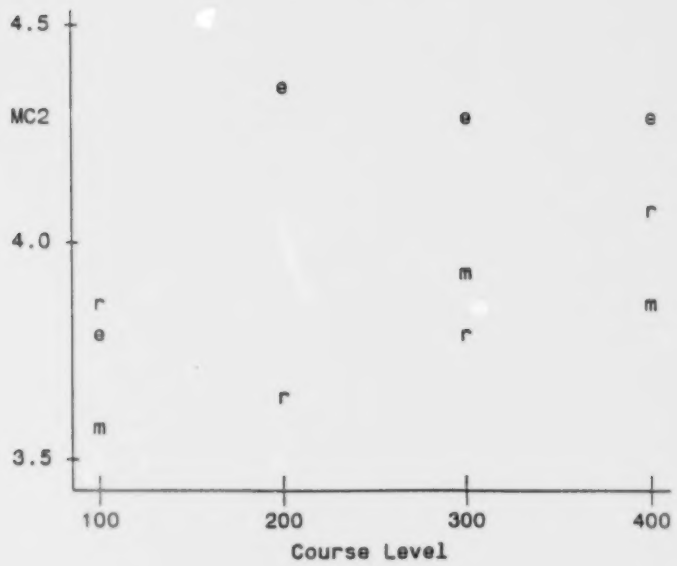


Figure 9. Scatter plots of the descriptive statistics for course motivation, course level and College of Liberal Arts and Science: AREA II. Symbol is the value of course motivation

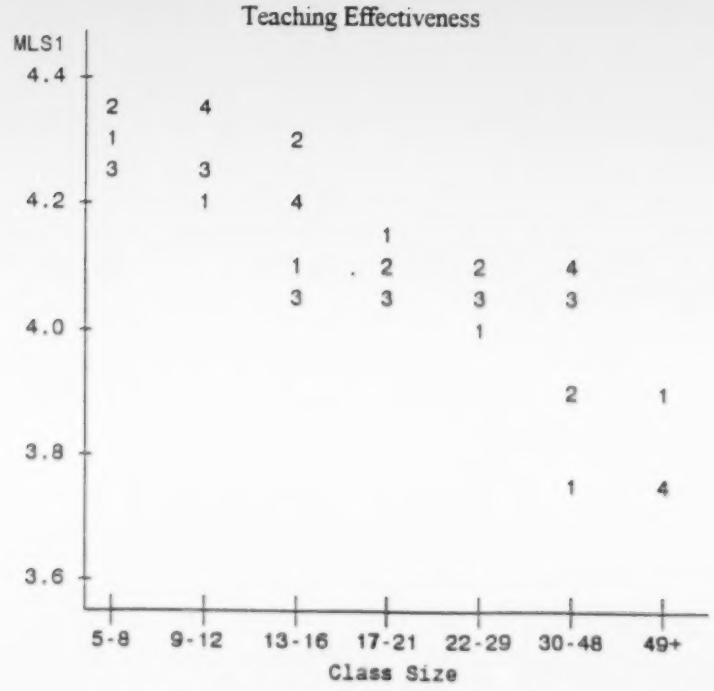
Table 8

Descriptive Statistics for Course Level and Class Size

Class Size	N	Mean	Std Dev	Min	Max
		Teaching Effectiveness			
Course level: 100					
5-8	50	4.31	.46	3.13	5.00
9-12	98	4.19	.63	1.58	5.00
13-16	131	4.11	.62	2.25	5.00
17-21	119	4.14	.62	1.71	5.00
22-29	80	3.98	.58	2.41	4.92
30-48	95	3.76	.73	1.92	4.91
49+	190	3.90	.59	1.79	4.96
Course level: 200					
5-8	81	4.36	.46	3.00	5.00
9-12	105	4.19	.57	2.18	5.00
13-16	110	4.28	.54	2.67	5.00
17-21	108	4.12	.47	2.85	4.95
22-29	156	4.10	.58	2.26	5.00
30-48	144	3.90	.62	1.97	4.96
49+	149	3.91	.66	1.94	4.99
Course level: 300					
5-8	175	4.26	.50	2.38	5.00
9-12	209	4.25	.53	2.22	5.00
13-16	197	4.07	.55	2.56	5.00
17-21	219	4.03	.61	1.70	5.00
22-29	224	4.03	.55	2.11	4.96
30-48	230	4.06	.55	2.13	4.94
49+	97	3.91	.66	1.79	4.88
Course level: 400					
5-8	267	4.37	.50	2.29	5.00
9-12	198	4.36	.46	2.45	5.00
13-16	101	4.19	.49	2.77	5.00
17-21	84	4.13	.55	2.20	5.00
22-29	61	4.06	.59	2.17	4.96
30-48	61	4.11	.54	2.38	4.95
49+	15	3.75	.70	2.77	4.86

Table 8 (Continue)

Class Size	N	Mean	Std Dev	Min	Max
		Course Quality			
Course level: 100					
5-8	49	4.23	.55	2.75	5.00
9-12	99	4.08	.65	1.92	5.00
13-16	132	4.11	.55	2.56	5.00
17-21	118	3.98	.58	1.95	5.00
22-29	80	3.86	.55	2.45	4.83
30-48	95	3.72	.62	2.30	4.83
49+	190	3.81	.45	2.56	4.92
Course level: 200					
5-8	81	4.35	.45	2.86	5.00
9-12	107	4.16	.48	2.67	4.92
13-16	107	4.21	.56	2.13	5.00
17-21	112	4.01	.50	2.50	4.95
22-29	149	4.05	.53	2.48	4.97
30-48	148	3.84	.53	2.36	4.80
49+	149	3.83	.58	2.03	4.95
Course level: 300					
5-8	176	4.23	.50	2.71	5.00
9-12	214	4.25	.47	2.56	5.00
13-16	190	4.02	54	2.43	5.00
17-21	221	3.98	.56	2.00	5.00
22-29	226	3.97	.50	2.22	4.91
30-48	227	3.97	.48	2.38	4.88
49+	98	3.86	.57	1.91	4.72
Course level: 400					
5-8	271	4.34	.50	2.57	5.00
9-12	199	4.33	.43	2.73	5.00
13-16	100	4.15	.49	2.62	5.00
17-21	82	4.12	.54	2.30	4.94
22-29	62	4.03	.54	2.17	4.86
30-48	61	4.03	.50	2.73	4.89
49+	15	3.70	.58	2.84	4.58



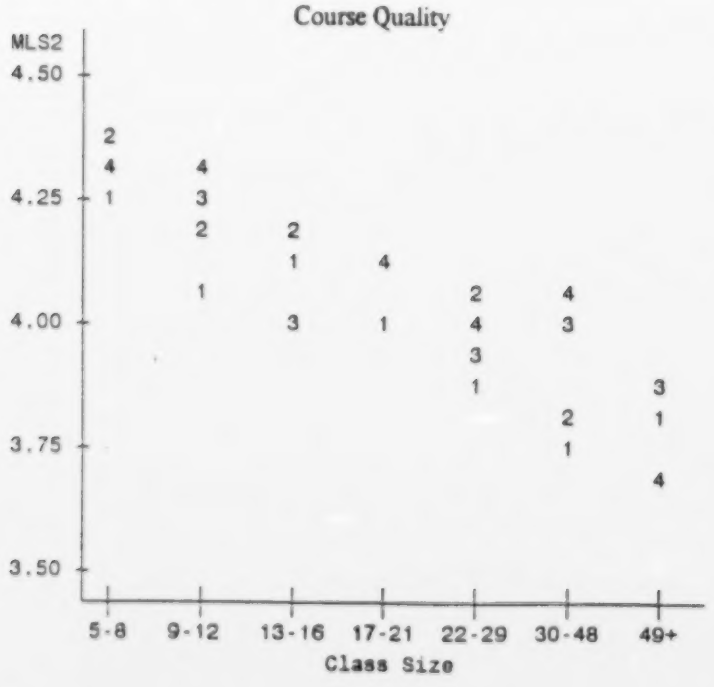


Figure 10. Scatter plots of the descriptive statistics for class size and course level. Symbol is the value of course level. The mls1 and mls2 are the average of the class ratings.

Table 9

Percentile for ICES Data (Fall 94 and Spring 95)

Class Size	N	Mean	StdDev	Low 10%	Middle 30%	Middle 70%	High 96%
			Teaching Effectiveness				
Course motivation: Required							
5-8	170	4.28	.52	3.33	3.89	4.67	5.00
9-12	212	4.16	.62	3.25	3.79	4.53	5.00
13-16	211	4.05	.64	3.15	3.68	4.42	4.95
17-21	237	4.04	.62	3.16	3.68	4.42	4.95
22-29	229	3.98	.60	3.10	3.62	4.34	4.86
30-48	260	3.88	.64	3.01	3.53	4.24	4.75
49+	253	3.82	.65	2.96	3.47	4.17	4.68
Course motivation: Mixed							
5-8	137	4.33	.45	3.55	4.01	4.65	5.00
9-12	156	4.29	.48	3.54	3.98	4.59	5.00
13-16	164	4.13	.53	3.40	3.83	4.42	4.85
17-21	147	4.13	.56	3.42	3.84	4.42	4 84
22-29	172	4.08	.56	3.38	3.79	4.36	4.78
30-48	182	4.05	.57	3.36	3.77	4.33	4.74
49+	139	3.93	.57	3.25	3.65	4.20	4.60
Course motivation: Elective							
5-8	266	4.36	.50	3.64	4.06	4.65	5.00
9-12	242	4.34	.48	3.66	4.06	4.62	5.00
13-16	164	4.29	.45	3.63	4 02	4.56	4.95
17-21	146	4.13	.52	3.48	3.87	4.40	4.78
22-29	120	4.12	.52	3.48	3.86	4.38	4.76
30-48	88	4.07	.57	3.44	3.81	4.33	4.70
49+	59	4.20	.57	3.58	3.95	4.45	4.82

Table 9 (Continue)

Class Size	N	Mean	StdDev	Low 10%	Middle 30%	Middle 70%	High 90%
			Course Quality				
Course motivation: Required							
5-8	175	4.23	.53	3.42	3.90	4.56	5.00
9-12	215	4.13	.56	3.37	3.82	4.45	4.90
13-16	201	4.00	.61	3.25	3.69	4.30	4.74
17-21	244	3.92	.57	3.19	3.62	4.21	4.64
22-29	224	3.87	.57	3.15	3.58	4.16	4.59
30-48	262	3.78	.54	3.07	3.49	4.07	4.48
49+	253	3.71	.52	3.02	3.43	4.00	4.41
Course motivation: Mixed							
5-8	135	4.29	.45	3.59	4.00	4.58	4.99
9-12	159	4.24	.48	3.59	3.98	4.51	4.90
13-16	165	4.08	.51	3.45	3.82	4.34	4.71
17-21	145	4.05	.53	3.44	3.80	4.31	4.67
22-29	171	4.03	.49	3.43	3.79	4.28	4.63
30-48	181	4.00	.49	3.41	3.76	4.24	4.59
49+	140	3.88	.48	3.31	3.65	4.12	4.46
Course motivation: Elective							
5-8	267	4.34	.50	3.67	4.07	4.62	5.00
9-12	245	4.31	.44	3.68	4.06	4.57	4.94
13-16	163	4.27	.44	3.66	4.02	4.52	4.88
17-21	144	4.12	.51	3.52	3.88	4.36	4.71
22-29	122	4.11	.45	3.53	3.87	4.35	4.69
30-48	88	4.05	.51	3.48	3.82	4.28	4.62
49+	59	4.14	.51	3.58	3.91	4.37	4.70